Structure, Vol. 13, 691–701, May, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.02.013

Processing of a 22 kDa Precursor Protein to Produce the Circular Protein Tricyclon A

Jason P. Mulvenna, Lillian Sando, and David J. Craik*
Institute for Molecular Bioscience
University of Queensland
Brisbane, Queensland, 4072
Australia

Summary

Cyclotides are a family of plant proteins that have the unusual combination of head-to-tail backbone cyclization and a cystine knot motif. They are exceptionally stable and show resistance to most chemical, physical, and enzymatic treatments. The structure of tricyclon A, a previously unreported cyclotide, is described here. In this structure, a loop that is disordered in other cyclotides forms a β sheet that protrudes from the globular core. This study indicates that the cyclotide fold is amenable to the introduction of a range of structural elements without affecting the cystine knot core of the protein, which is essential for the stability of the cyclotides. Tricyclon A does not possess a hydrophobic patch, typical of other cyclotides, and has minimal hemolytic activity, making it suitable for pharmaceutical applications. The 22 kDa precursor protein of tricyclon A was identified and provides clues to the processing of these fascinating miniproteins.

Introduction

Circular proteins have been discovered in bacteria, plants, and animals over recent years (Trabi and Craik, 2002). The cyclotides (Craik et al., 1999) are a group of plant-derived proteins that combine backbone cyclization with an embedded cystine knot in their unique three-dimensional structure. This family, comprised of two subfamilies termed the Bracelet and the Möbius (Craik et al., 1999), has attracted great interest because their characteristic structural motif, dubbed the cyclic cystine knot (CCK), imparts considerable stability to these molecules. Cyclotides are resistant to a range of adverse thermal, chemical, and enzymatic conditions that would be intolerable for other proteins (Colgrave and Craik, 2004) and possess a range of interesting bioactivities (Craik et al., 2004), including anti-HIV activity (Daly et al., 1999a). The range of activities and the unusual stability of the framework make the CCK motif an exciting prospect as a structural framework for the development of stable peptidic drugs (Craik et al., 2002).

Recent studies of the cyclotides have elucidated their unusual three-dimensional structure. Figure 1 shows a schematic representation of the general structural framework as deduced from the structures reported to date (Barry et al., 2004; Craik et al., 1999; Daly et al., 1999a; Saether et al., 1995; Trabi and Craik, 2004). At the core of the framework is the cystine knot (Craik et al., 2001), comprised of the ring formed by Cys I–IV and II–V and the connecting backbone residues, which is penetrated by the third disulfide formed between Cys III and VI. Conceptually, the cyclotides can be thought of as a series of six loops between successive Cys residues that constitute the knotted scaffold. Loops 1 and 4 form part of the embedded ring, while loops 2, 3, 5, and 6 are solvent exposed and connect the edges of the knot. Significantly, the loops displaying the greatest conservation are those forming the cystine knot itself —loops 1 and 4—which are highly conserved in length as well as amino acid composition. Loops not directly involved in the cystine knot exhibit a much greater amount of variation. This pattern is repeated on a structural level, with the cystine knot motif conserved and the pendant loops exhibiting variable structural characteristics. The combination of a conserved CCK motif with variable loop regions suggests that the CCK framework may function in the plant as a stable combinatorial scaffold for producing proteins of varying biological activity.

Backbone cyclization has recently attracted interest as a mechanism for increasing the stability of proteins both thermodynamically and by providing increased resistance to protease activity (Camarero et al., 2001; Deechongkit and Kelly, 2002). As one of the few eukaryotic macrocyclic gene products, the cyclotides are the result of a naturally evolved backbone cyclization mechanism. Determination of the mechanism of cyclization would hence be of great utility in the development of methodologies for the stabilization of protein drugs via backbone cyclization. Analysis of mRNA transcripts in the plants *Oldenlandia affinis* and *Viola odorata* showed that cyclotides are initially expressed as linear precursors (Dutton et al., 2004; Jennings et al., 2001). Figure 1 shows the organization of two of these precursors. Each contains a signal sequence followed by a precursor sequence and one, two, or three copies of mature cyclotide sequence. In multicopy transcripts, the C-terminal end of the precursor sequence is repeated N-terminally to each mature sequence. To differentiate the two regions, the repeated sequence has been termed the N-terminal repeat (NTR) and the nonrepeated segment of the precursor the proregion. A short hydrophobic tail is found at the end of each precursor. The processing of this linear product into a cyclic protein requires two cleavage reactions and a cyclization reaction. While the latter is poorly understood, the cleavage reactions occur N-terminally after a Lys/Gly/Asn residue and C-terminally after an Asn/Asp residue in loop 6 (Dutton et al., 2004; Jennings et al., 2001). The mechanism by which the cyclotide is consequently cyclized is unknown, although the conservation of an Asn/Asp residue in loop 6 in all but one known cyclotide suggests that this residue may play an important role in the mechanism.

Although inspection of mature cyclotide sequences provides clues about the cyclization mechanism, it is

*Correspondence: d.craik@imb.uq.edu.au

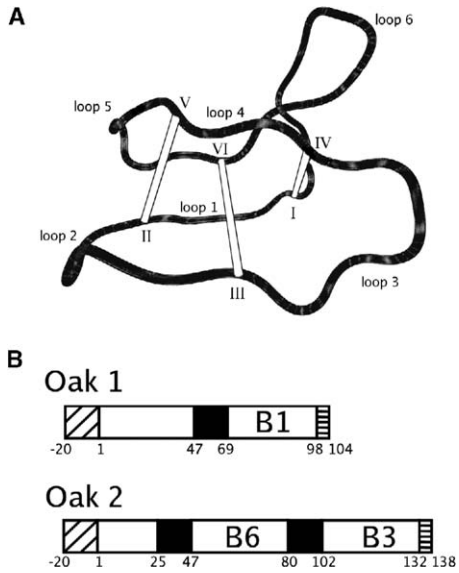

Figure 1. Generalized Structure of a Cyclotide and Linear Precursors Identified from *O. affinis*

(A) The general structure and loop nomenclature of the cyclotides. The cyclotide framework consists of an embedded cystine knot made up of loops 1 and 4 and three disulfide bridges between Cys I-IV, II-V, and III-VI. The remainder of the protein consists of loops 2, 3, 5, and 6, which connect successive Cys residues. Cys residues are numbered in roman numerals according to the order in which they appear in the linear precursor.

(B) Two of the four kalata precursors that have been identified in *O. affinis* (Jennings et al., 2001). The precursors consist of a signal sequence (diagonal hash), a short proregion (gray), and between one and three mature protein sequences, each preceded by an NTR sequence (black). Each precursor has a short hydrophobic tail (horizontal hash). Examples are shown of precursors containing one (B1) or two (B3 and B6) mature protein sequences. *Oak4* (Jennings et al., 2001) contains three copies of the cyclotide kalata B2.

likely that elements C-terminal to the mature sequence also play an important role. Accordingly, as part of our discovery program, we have been searching for "unusual" cyclotides to be characterized at both a structural and a genetic level. In this paper, we report the discovery and characterization of a novel, to our knowledge, cyclotide precursor as well as the three-dimensional structure of the mature protein. This cyclotide, named tricyclon A, is 33 amino acids in length and is of great interest because, not only does it comprise elements of both cyclotide subfamilies, but it possesses the largest and most sequentially diverse loop 6 yet discovered in the cyclotides. Because loop 6 is formed when linear precursors are transformed into cyclic proteins, tricyclon A not only affords an opportunity to study the structural ramifications of increased loop size on cyclotide structure, but it may also yield important clues as to the cyclizing mechanism of these fascinating molecules.

Results

Isolation and Sequence Determination

As part of a screening program to identify cyclotides, flowers from *Viola tricolor* were analyzed by using a combination of RP-HPLC and mass spectrometry. Cyclotides typically elute late in RP-HPLC, and preliminary experiments revealed several late-eluting peaks, one of which contained a constituent with an average mass of 3480 Da. Large-scale extraction of 40 g fresh *V. tricolor* flowers produced approximately 600 mg of dry plant extract. RP-HPLC of this dried extract yielded approximately 3 mg of the purified protein.

For sequencing approximately 1 mg of protein was reduced with DDT and then alkylated with iodoacetamide, yielding a molecule 348 Da heavier than the native protein, indicating the addition of six alkyl groups to six Cys residues. The sample was split in two, and each sample was digested with either trypsin or Endo Glu-C. Analytical RP-HPLC of the Endo Glu-C digest revealed a peak containing a constituent with a mass of 3846 Da. This mass was equivalent to the mass of tricyclon A (3480 Da), with six alkyl groups (348 Da) and a single water molecule. This fragment was N-terminally sequenced. The tryptic digest produced a complex mixture of products, and these were sequenced by using tandem mass spectrometry. One fragment produced a sequence that overlapped considerably with the Edman sequencing, and the combination of these two sequences gave a mass equivalent to 3480 Da if a cyclic structure was assumed and 18 Da, the mass of a single water molecule, was removed from the linear sequence. Because Leu and Ile possess the same mass, the identity of residue 31 could not be unambiguously defined by using MS-MS, and it was only when the gene sequence was obtained that the residue was unambiguously determined to be Ile. The use of two different proteases to produce sequence data also confirmed the cyclic nature of tricyclon A, as a combination of the two sequences gave an uninterrupted run of sequence without an N or C termini. The complete sequence is given in Figure 2.

Concurrent screening in the plant *V. arvensis* led to the discovery of a protein with the same mass and retention time on RP-HPLC as tricyclon A. This is consistent with the distribution pattern of other cyclotides, including kalata B1, which are often found in a variety of plants.

^{1}H-NMR Resonance Assignments

The resonances in the two-dimensional TOCSY spectrum were well dispersed in the amide region, with only overlap between Cys17 and Gly18 complicating identification of the individual amino spin systems. Analysis of primary NMR data suggests that the major feature of tricyclon A is a triple-stranded, anti-parallel β sheet between residues 16 and 32 that contains two turns centered on residues 18–21 and 26–29. Figure 3 summarizes this secondary structure. The central strand of the sheet is formed by residues 22–27, and 4 of these residues have slowly exchanging amide protons, as would be expected in a β sheet. All other amide protons that are predicted to make hydrogen bonds across the β sheet are slow exchanging and, apart from two cases in which overlap prevented unambiguous assignment, NOE connectivities are consistent with the secondary structure. The amide proton for Cys15 is also slow exchanging but is not predicted to participate in the hy-

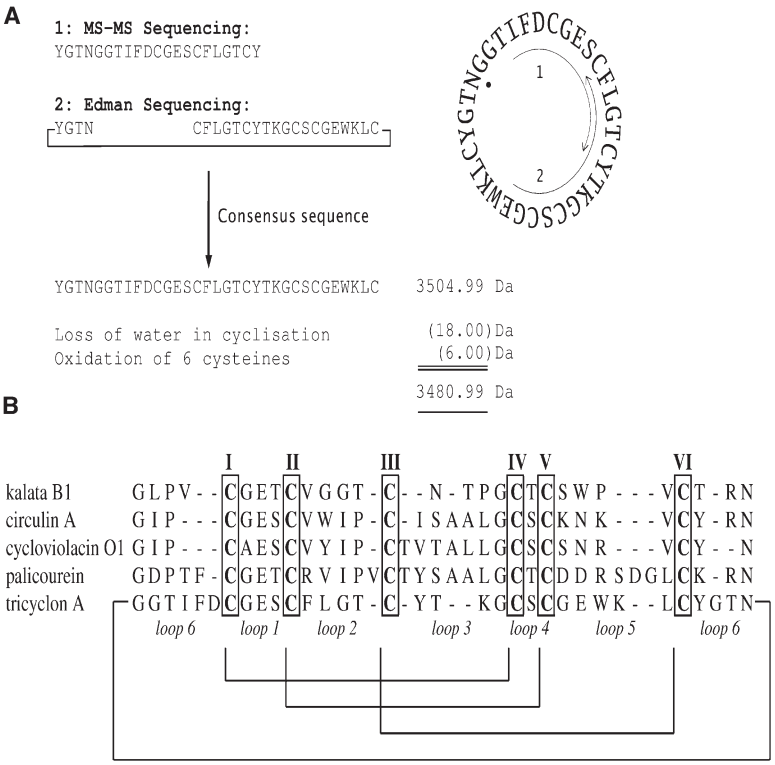

Figure 2. Sequence of Tricyclon

(A) Sequence of tricyclon A showing results from Edman and MS-MS sequencing and the mass calculations used to confirm the sequence. The overlapping of the respective sequences highlights the cyclic nature of the protein.

(B) Sequence alignment of five representative cyclotides showing the conserved Cys framework and the extended sequence of loop 6 in tricyclon A. The absolutely conserved Cys residues that form the cystine knot are boxed, and the disulfide bonds are represented by lines joining the respective Cys residues. It can be seen that in loop 6 tricyclon A contains the highly conserved Asn residue but lacks the equally well conserved Pro residue. In addition, the size of loop 6 in tricyclon A is two residues greater than that of palicourein, the largest cyclotide yet characterized.

drogen bonding of the β sheet and is likely involved in hydrogen bonding elsewhere in the structure. Although a small amount of CD_3CN was added to the sample to increase its solubility, it is unlikely that the cosolvent assists in stabilizing the secondary structure reported here.

A recent report confirmed the disulfide connectivity of the cyclotides kalata B1 and cycloviolacin O1 by analysis of the χ^1 angles of the six Cys residues (Rosengren et al., 2003). Similarly, the disulfide connectivity of tricyclon A was confirmed by using χ^1 dihedral angles. The χ^1 angles for each of the Cys residues were determined by using a combination of $^3J_{H\alpha H\beta}$ coupling constants and interproton HN-Hβ distances. For cysteines 1, 10, 15, and 23, an angle of –60° was derived, while Cys5 was determined to possess a χ^1 dihedral angle of +60° and Cys17 a χ^1 angle of +180°. These angles are consistent with the χ^1 dihedral angles reported for kalata B1 and cycloviolacin O1 (Rosengren et al., 2003) and would indicate that tricyclon A possesses the same knotted disulfide connectivity.

Three-Dimensional Structure

A set of 50 structures was calculated by simulated annealing with 194 interresidue NOE distance restraints, including 75 long-range (between residues separated by more than 3 amino acids) NOEs, and 29 dihedral angle restraints. Figure 4 shows a backbone superimposition of the 20 lowest-energy structures after refinement in a water box that illustrates that most regions of the structure are well defined. The mean pairwise rmsd over the whole molecule was 0.65 Å for the back-

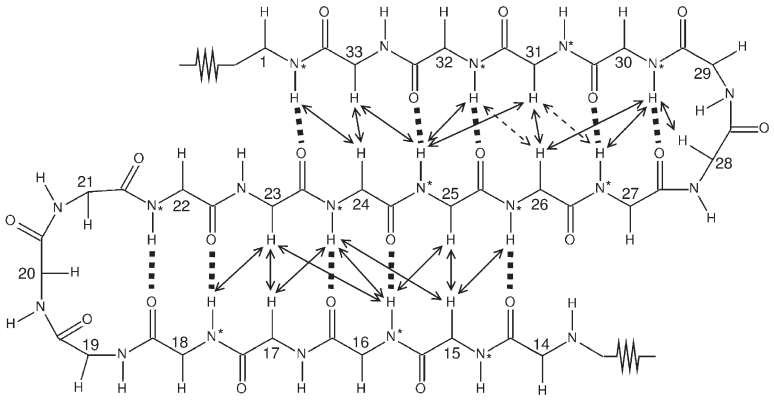

Figure 3. Secondary Structure of Tricyclon A

A schematic representation of the β sheet region of tricyclon A. Interstrand NOEs are shown as arrows, slowly exchanging amide protons (after dissolution in ^2H_2O) are indicated with a asterisks, potential hydrogen bonds are depicted with broken dashes, and ambiguous NOEs that could not be resolved due to overlap are represented by dashed arrows. Sequential NOEs are not shown for clarity.

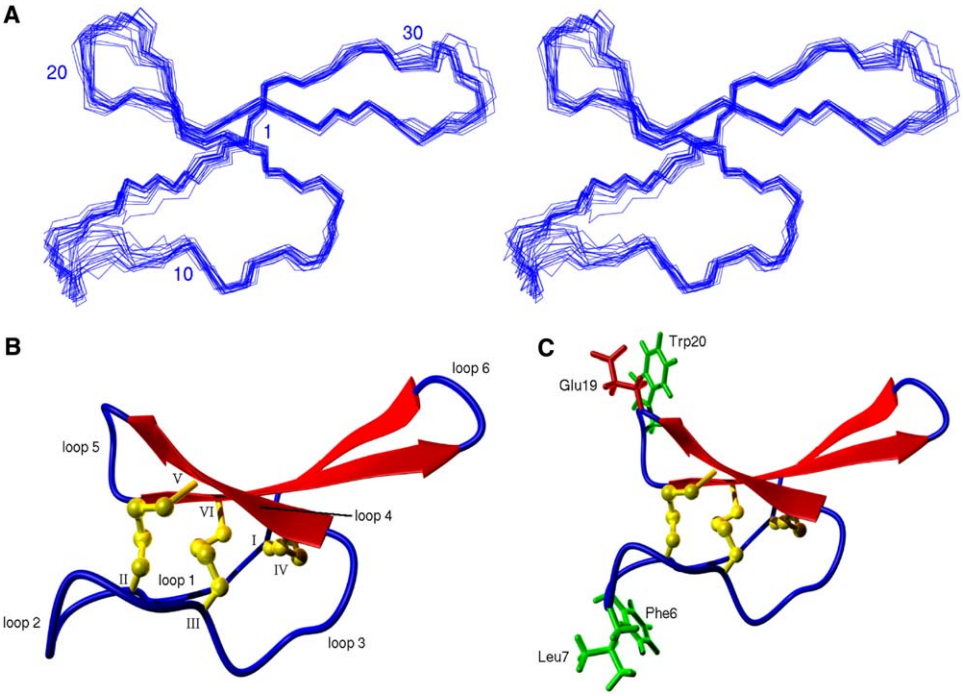

Figure 4. Structure of Tricyclon A

(A) Stereoview of the 20 lowest-energy structures generated after minimization in a water shell. Every tenth residue is numbered.

(B) Ribbon diagram showing the structure of tricyclon A, which includes the cystine knot (Cys residues are numbered with roman numerals) and the triple-stranded β sheet.

(C) Ribbon diagram highlighting the side chains of Trp20, Glu19, Phe6, and Leu7. In other cyclotides, hydrophobic interactions between corresponding residues are thought to stabilize the β sheet, but in tricyclon A, no interactions are apparent. The stereo and ribbon views were produced by using the program MOLMOL (Koradi et al., 1996), and ribbon diagrams were further processed by using the program POVRAY.

bone atoms and 1.60 Å for all heavy atoms. A summary of the structural statistics is given in Table 1. The structures show no major divergence from ideal covalent geometry and fit the experimental restraints with minimal violations. No NOE violations greater than 0.22 Å or dihedral angle restraint violations greater than 2.20° were present in the set of lowest-energy structures.

Analysis of the structures with the programs PROMOTIF (Hutchinson and Thornton, 1996), MOLMOL (Koradi et al., 1996), and InsightII shows that the main element of secondary structure is a three-stranded, anti-parallel β sheet. This β sheet consists of two hairpins within residues 15–27 and 22–32. An inverse γ turn is located in loop 2, directly beneath loop 5. The existence of a formally recognized β sheet is unusual among the cyclotides, as the third strand that includes residues from loop 1 is normally distorted. In this case, the β sheet is formed entirely from residues in loops 3, 4, and 6, giving loop 6 a more structured nature than in previously characterized cyclotides. Given that the spectra for tricyclon A are no better resolved than other cyclotide structures, the absence of a β sheet in loop 6 of other cyclotides is not due to the collection of better distance restraints in tricyclon A. A type II β turn is also found in loop 3 of the protein, which connects Cys10 to Cys15. The amide proton of Cys15 exchanges slowly, and examination of the structures suggests that it interacts with the carboxyl oxygen of Thr12, stabilizing this β turn.

Although not confirmed experimentally, it has been suggested that interactions observed between the hydrophobic residues of loop 2 and loop 5 stabilize the distorted β sheet in kalata B1 and cycloviolacin O1 (Rosengren et al., 2003). Figure 4 shows that the corresponding residues in tricyclon A are orientated away from each other, suggesting that these interactions are not present. The Glu residue in loop 1 of the cyclotides is, apart from the six Cys residues, one of the few residues absolutely conserved throughout all known cyclotides. In kalata B1 and cycloviolacin O1, the oxygen atoms of the Glu3 carboxyl group forms hydrogen bonds with the backbone amides of residues in loop 3 (residues 11, 12, and 13 in cycloviolacin and residues 11 and 12 in kalata B1) as well as the hydroxyl proton of a Thr residue (residue 13 in kalata B1 and residue 13 in cycloviolacin O1) (Rosengren et al., 2003). It can be seen in Figure 5 that Glu3 of tricyclon A is oriented such that similar interactions may occur. This is further supported by shifts of the amide protons of residues 11 and 12 during pH titrations of tricyclon A; however, confirmation of the interaction was hampered by severe spectral broadening at intermediate pH values that prevented calculation of a pKa value. Overall, the structure for tricyclon A shows a fold similar to that adopted by

Table 1. Geometric and Energetic Statistics for Tricyclon A

Energies (kcal mol^{-1})	
Overall	–1067.22 ± 16.47
Bonds	7.30 ± 0.79
Angles	41.38 ± 4.53
Improper	7.95 ± 0.96
Van der Waals	–80.55 ± 7.42
NOE	20.18 ± 3.94
cDih	0.41 ± 0.26
Dihedral	156.63 ± 7.02
Electrostatic	–1220.51 ± 21.21
Rmsd (Å)	
Bond	0.004 ± 0.0002
Angle	0.58 ± 0.03
Improper	0.45 ± 0.03
NOE	0.045 ± 0.005
cDih	0.46 ± 0.16
Pairwise rmsd (Å)	
Backbone	0.65 ± 0.19
Heavy	1.60 ± 0.22
Experimental Data	
Distance restraints	194
Dihedral restraints	30
NOE violations exceeding 0.20 Å	2 (largest 0.220)
CDih violations exceeding 2.0°	5 (largest 2.20)
Ramachandran	
Most favored	81.0%
Additionally allowed	19.0%
Disallowed	0.0%

The values are given as mean ± standard deviation for the ensemble of the 20 final solution structures. Experimental distance restraints include only interresidual NOEs.

other characterized cyclotides, but the presence of a formally recognized β sheet incorporating loop 6 makes it of particular interest.

RT-PCR and RACE

On the basis of the suggestion that the processing point of the cyclotides is in loop 6 (Jennings et al., 2001), a primer was designed for loop 2 of the mature protein sequence for use with oligo-dT in RT-PCR. An approximately 350 bp fragment was produced in this reaction and was found to contain the complete coding sequence for a cyclotide similar to tricyclon A, denoted tricyclon B, as well as the partial sequence of tricyclon A. A reverse primer was designed from this sequence for use in 5′-RACE. After the PCR reaction, a fragment of approximately 500 bp was obtained. This band contained the remainder of the sequence for tricyclon A as well as a signal sequence and precursor region.

Figure 6 sets out the sequence of the tricyclon A precursor and its structure. The organization of the gene is similar to that of other cyclotide cDNAs (Jennings et al., 2001) in that it contains a typical endoplasmic reticulum (ER) signal sequence of 22 amino acids and a precursor region of 49–50 amino acids that is repeated at the N-terminal region of each mature protein sequence. In the *Oak* clones from *O. affinis*, multicopy sequences are separated by an NTR domain, but a short proregion follows the ER signal before the initial NTR. In the case of tricyclon A, the entire precursor region constitutes the NTR and is repeated between the two mature protein sequences. A 15 amino acid region follows the tricyclon A mature protein sequence, and a four amino acid tail follows the tricyclon B sequence. The entire predicted precursor protein is 204 amino acids and has a predicted mass of approximately 22 kDa. This is substantially larger than any clones previously reported for cyclotides.

The three-dimensional structures of the NTR from *O. affinis* and *V. tricolor* have recently been determined (Dutton et al., 2004) and in both cases adopt an amphipatic helix. To determine whether the NTR region discovered here also adopts a helical conformation, the sequence was analyzed by using secondary structure prediction software. Figure 6 shows that the NTR has two regions of predicted helix, and given the conservation of a helical structure across two species from the phylogenetically diverse Rubiaceae and Violaceae plant

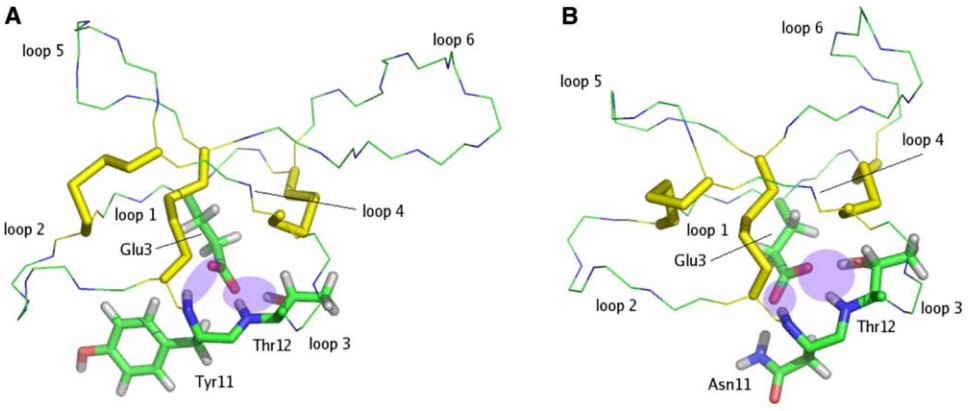

Figure 5. Hydrogen Bond Network Formed by Conserved Glu3

(A and B) Orientation of conserved Glu residue and the network of hydrogen bonds formed in (A) tricyclon A and (B) kalata B1. Atoms involved in hydrogen bonding have been indicated with a shaded circle, and for the stick representations, oxygen atoms are colored red, hydrogen atoms gray, nitrogen atoms blue, and carbon atoms green. The conservation of Glu3 across the cyclotides would appear to indicate that these hydrogen bond interactions are crucial for the CCK fold. Ribbon diagrams were produced by using the program PyMOL (DeLano, 2002).

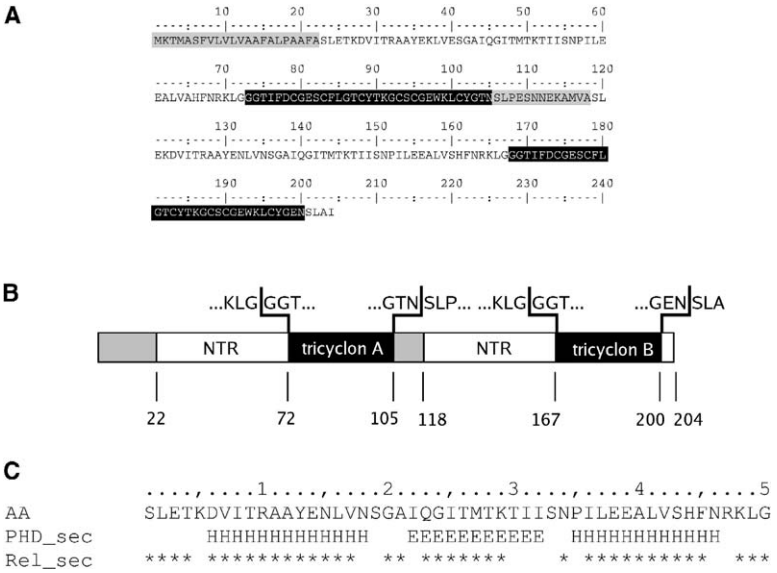

Figure 6. Organization and Sequence of Tricyclon A Precursor

(A) Sequence of the tricyclon A precursor. The sequence has a short ER signal sequence (shaded gray), a short proregion (white), and the mature sequence of tricyclon A (highlighted in black). Following this is a short tail region (also shaded gray) preceding a relatively long NTR region and the mature sequence of a cyclotide similar to tricyclon A, named tricyclon B (highlighted in black). A four residue tail follows the mature tricyclon B sequence.

(B) A schematic of the structure of the multitranscript gene showing the domains shaded in the same way as in (A).

(C) Helical regions in the NTR as predicted by the program PHDsec (Rost and Sander, 1993). "H" denotes helix, "E" denotes an extended structure, and an asterisk denotes a strong prediction.

families, it is very likely that the NTR discovered here also adopts a helical structure.

Hemolytic Assays

A number of cyclotides have been found to possess hemolytic activity (Barry et al., 2003; Daly et al., 1999a; Gustafson et al., 1994; Schopke et al., 1993; Tam et al., 1999; Witherup et al., 1994), and this activity is a potential obstacle when considering the cyclotides as drug leads. Hence, the hemolytic activity of tricyclon A was tested to determine whether it shared this characteristic with other cyclotides. In hemolytic assays conducted with 500 μM tricyclon A, only 2% lysis of RBCs was observed after 1 hr of incubation at 37°C. This can be contrasted with kalata B1, which produced 50% hemolysis at approximately 100 μM under the same conditions. Various values have been reported for the HD50 (dose that gives 50% hemolysis) of kalata B1 (Barry et al., 2003; Daly et al., 1999b; Tam et al., 1999) under different experimental conditions. For example, the reported value of 300 μM was based on assays carried out at room temperature (Barry et al., 2003). With our current assay conditions at 37°C, which are more physiologically relevant, the HD50 for kalata B1 is 100 μM. Under these conditions, tricyclon A has minimal hemolytic activity.

Discussion

The structure determined here for the macrocyclic protein tricyclon A is well defined and confirms that it belongs to the cyclotide class of proteins. Unlike other cyclotides that have been structurally characterized, tricyclon A contains a formally recognized triple-stranded β sheet that introduces a high degree of ordered structure to loop 6 of the CCK motif. The changes in loop 6 occur in a region implicated in the processing and cyclization of cyclotides from their precursor proteins, and hence the structure determined here provides additional insight into the cyclization mechanism. Furthermore, given recent interest in using small, stable proteins for scaffolds in the engineering of bioactive compounds (Craik et al., 2002; Martin et al., 2003), the existence of a well-defined structural element in a solvent-exposed loop indicates that the CCK framework is amenable to structural variation and offers an opportunity for the engineering of a wide range of structural functionalities into this highly stable motif.

Comparison to Other Cyclotide Structures

To date, five cyclotide structures have been determined: kalata B1 (Rosengren et al., 2003; Saether et al., 1995), circulin A (Daly et al., 1999a), cycloviolacin O1 (Craik et al., 1999; Rosengren et al., 2003), palicourein (Barry et al., 2004), and Vhr1 (Trabi and Craik, 2004). All exhibit a similar structure in which the backbone is folded back upon itself about the cystine knot core. In all cases, β strands forming a β hairpin are present, with a third strand inferred by NOE, coupling, and hydrogen bonding patterns. However, the third strand is not formally recognized by structure interrogation programs such as PROMOTIF, MOLMOL, or InsightII. Superimposition of the tricyclon A structure with kalata B1, circulin A, cycloviolacin O1, and palicourein shows an rmsd over the Cα of the Cys residues that comprise the cystine knot of 0.41, 0.55, 0.37, and 0.44 Å, respectively. This clearly indicates that the core of tricyclon A corresponds closely to that of the previously characterized cyclotides. The side chain χ^1 angles of the Cys residues reported here also confirm the structural similarity of the cystine knot core, as they are similar to those derived for kalata B1. The angles also confirm the disulfide connectivity of tricyclon A, as they are consistent with the knotted topology as previously reported (Rosengren et al., 2003).

Apart from the cystine knot core, tricyclon A possesses further similarities with other cyclotide structures. In particular, Gly14 has a positive ϕ angle, which is consistent with the structures of kalata B1 and cycloviolacin O1. A Gly is highly conserved in the last posi-

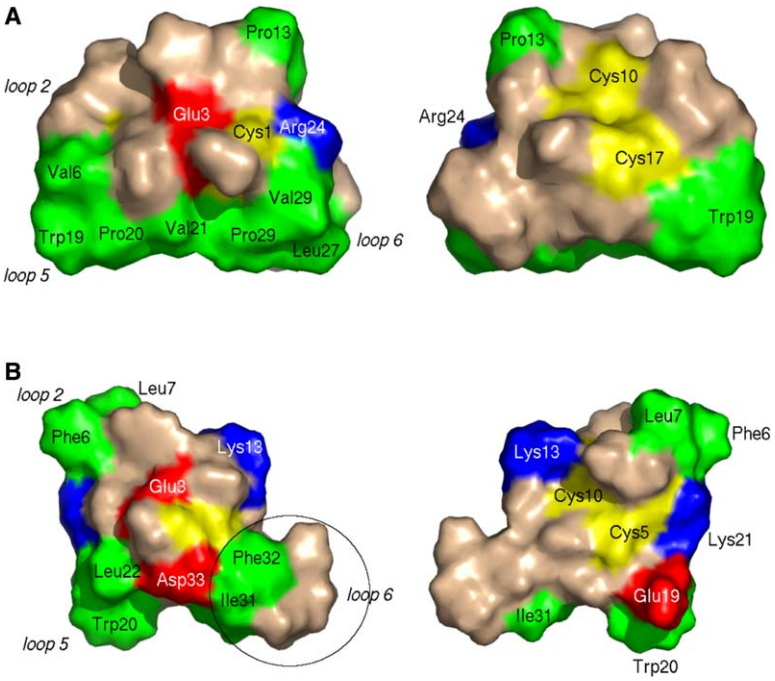

Figure 7. Surface Comparison of Kalata B1 and Tricyclon A

(A) Surface rendering of the prototypic cyclotide kalata B1 with hydrophobic residues colored green, positively charged residues colored blue, and negatively charged residues colored red. For orientation, loops 2, 5, and 6 are labeled on the left-hand figure.

(B) Surface rendering of tricyclon A colored and labeled similarly to kalata B1. The protruding β hairpin formed by loop 6 of tricyclon A is circled in the left-hand figure. In both panels, the right-hand view is rotated 180° about the y axis. Surface renderings were produced with the program PyMOL (DeLano, 2002).

tion of loop 3 within the cyclotides, and the positive ϕ angle is thought to be necessary for linking loop 3 to the cystine knot. As shown in Figure 5, the structure reported here places the oxygen atoms of the Glu3 side chain in close proximity to residues in loop 3—namely Tyr11 and Thr12—suggesting that hydrogen bond interactions similar to those reported for kalata B1 and cycloviolacin O1 (Rosengren et al., 2003) are also present in tricyclon A. This is further supported by the shifts in the amide protons of these residues during pH titrations. Apart from the six Cys residues, Glu3 is the only residue completely conserved within the cyclotide family; hence, these interactions may be crucial for the CCK fold.

Despite these similarities, the structure of tricyclon A contains key differences, and these differences illustrate the potential for structural variation inherent within the CCK motif. The most striking difference between tricyclon A and other cyclotide structures is the triple-stranded β sheet made up of residues from loops 4 and 6. This is a classical β sheet that is formally recognized by PROMOTIF, MOLMOL, and InsightII. The sheet incorporates all the residues from loop 6, which distinguishes tricyclon A from the other cyclotides in which loop 6 adopts an extended fold. Figure 7 shows a comparison of the surfaces of kalata B1 and tricyclon A, and it can be seen that the β hairpin of loop 6 protrudes from the globular core of tricyclon A. This contrasts with other cyclotides that possess a more globular structure. It is interesting to note that β hairpins have been identified as common motifs in small anti-microbial proteins (Bulet et al., 2004) and are often important for protein-protein interactions (Dou et al., 2004) and active site presentation. At least one example of an insecticidal protein that relies on an exposed hairpin is the spider toxin ω-atracotoxin-Hv1a (Tedford et al., 2001). In this protein, the putative active site is situated on a β hairpin that protrudes from the globular, disulfide-rich core in a similar fashion to tricyclon A. It has been suggested that the cyclotides have a role in plant defense (Craik, 2001), and by analogy to ω-atracotoxin-Hv1a, it is possible that the protruding hairpin in this structure may have an important role in the natural function of tricyclon A.

The structure and activity results presented here for tricyclon A suggest that the cyclotides can be engineered to reduce hemolytic activity while retaining the structural integrity of the CCK motif. Although it has been suggested that hydrophobic interactions stabilize the distorted β sheet in other cyclotides (Rosengren et al., 2003), tricyclon A does not possess a hydrophobic patch, with the surface consisting of hydrophobic residues interspersed with charged or polar residues (Figure 7), and, as discussed elsewhere, hydrophobic interactions between loops 2 and 5 appear to be absent in tricyclon A. The lack of a hydrophobic patch may explain the low hemolytic activity of tricyclon A, which distinguishes it from the other characterized cyclotides, and the absence of interactions between loops 2 and 5 suggest that surface hydrophobicity is not needed to stabilize the CCK motif. In this respect, tricyclon A provides an example of the way that the CCK motif can be adapted to include, or remove, structural motifs that may provide specific bioactivities.

The cyclotides have been divided into two subfamilies, the Bracelet and Möbius, on the basis of a cis-Pro peptide bond in loop 5 of Möbius family members (Craik et al., 1999). Loop 2 of tricyclon A bears a closer resemblance to the Möbius subfamily, while loop 5 is similar to the corresponding loop in the Bracelet subfamily. Not only does this make tricyclon A a rare hybrid of the Bracelet and Möbius subfamilies, but in each

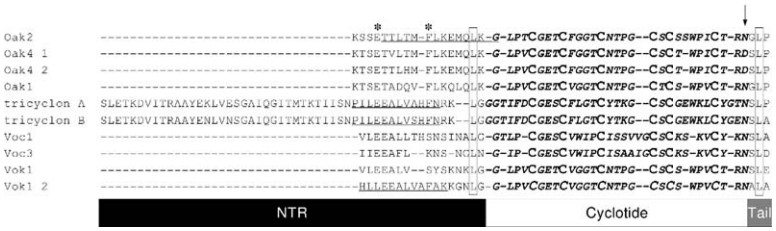

Figure 8. Comparison of the NTR from *O. affinis*, *V. odorata*, *and V. tricolor*

Alignment of a selection of the repeated domains, containing the NTR and mature cyclotide sequence, identified in cyclotide precursors. The C-terminal processing point has been marked with an arrow, and the absolutely conserved Leu residues situated one residue away from both the N- and C-terminal processing points have been boxed. The absolutely conserved Glu residue three or four residues away from the N terminus of the NTR has been marked with an asterisk, as has the partially conserved Phe residue. The NTR sequences that have been structurally characterized as helical in *Oak2* and *Vok1* have been underlined, and, similarly, the corresponding regions of predicted helices in the tricyclon repeated domains have also been underlined. Cys residues in the mature cyclotide sequence have been highlighted in larger font, and the mature cyclotide sequence has been italicized. Inclusion of multiple repeated domains from the same precursor has been indicated by numbering after the precursor name. The schematic below the alignment shows the nomenclature of the different regions contained within the repeated domain.

case, the hybrid loop "choice" has resulted in the removal of a Pro that is conserved in the respective subfamily. In addition, the almost universally conserved Pro in loop 6, which forms a part of the hydrophobic patch in each subfamily, is not present in tricyclon A. Given the interest in the evolution of the cyclotides and the rarity of hybrids, an example exhibiting sequence similarities to both subfamilies is of great interest.

Cyclotides and Cyclization

One of the challenges in the field of cyclic proteins is that cyclization leaves no "footprint" and determining the end points of a linear precursor is not possible if only the mature cyclic sequence is known. The cyclizing points of the prototypic cyclotide kalata B1 were only determined after the isolation of mRNA coding for the linear precursor (Jennings et al., 2001). In this case, the C-terminal processing point was shown to be after the conserved Asn residue in loop 6, with four N-terminal residues immediately preceding CysI incorporated into the mature protein. A subsequent study showed the same processing points for cyclotides from *V. odorata* (Dutton et al., 2004). Figure 8 shows an alignment of the repeated domains, containing the NTR and the mature cyclotide, that have been identified in other cyclotide precursors with the repeated domains reported here. The sequences of tricyclon A and B confirm the conservation of the C-terminal processing point as an Asn/Asp residue, but the Leu(Ile)/Pro motif at the N-terminal part of the mature protein is not conserved. This suggests that this motif, conserved in all other cyclotide precursors (Dutton et al., 2004; Jennings et al., 2001), is not important for cyclotide processing and may serve some other functional or structural role in the cyclotides that possess it.

Interestingly, there is a complete conservation of a Leu one residue away from the cleavage points at both ends of the mature sequence. Leu and Ile residues have been shown to be involved in vacuolar targeting in other proteins (Brown et al., 2003; Frigerio et al., 2001) and, for example, are essential in the N-terminal propeptide containing the sequence NPIRL, or variations, for sorting of proteins to lytic vesicles in tobacco (Matsuoka and Nakamura, 1999). Interestingly, the hydrophobic tail found in the cyclotide precursors could be considered to be a member of the C-terminal vacuolar signal sequences (ctVSS) found in barley lectins (Dombrowski et al., 1993), bean phaseolin (Frigerio et al., 1998), brazil nut 2S albumin (Saalbach et al., 1996), and tobacco chitinase A (Neuhaus et al., 1994) that reside on proteins destined for protein storage vesicles (PSV). The ctVSS signals have no standard length or sequence conservation but possess only a hydrophobic character as a distinguishing feature. The strong conservation of Asn/Asp at the C-terminal processing point may also indicate the importance of an Asn-specific proteinase that processes several protein precursors, including the lectins, proteolytic enzymes, and the storage proteins 2S albumins and 11S globulin; in many cases this processing also occurs in a PSV (Muntz and Shutov, 2002). Whether an Asn-specific proteinase is involved in mechanisms by which the N-terminal cleavage of the precursor and cyclization of mature cyclotide are achieved remains a mystery. An alternative consideration could be that the Leu residues act as recognition sites for proteolysis of the precursor into separate units with extraneous amino acids (one residue N-terminally and two C-terminally) removed during an as yet unidentified cyclization reaction.

It can also be seen in Figure 8 that the NTR domains are highly homologous within species, but between species the sequences vary greatly. Although almost nothing is known about the folding and processing of these precursors, it is interesting that a conserved element of the precursor should appear to tolerate such sequence variation. Nonetheless, despite the low sequence identity, a Glu residue is conserved three or four residues from the N terminus of the NTR in all precursors characterized, and closer to the C terminus of the NTR a Phe is conserved in all of the *Oak* clones. These residues are also present in the sequences reported here, and they may have some functional importance for the processing of the precursor.

The organization of the tricyclon A precursor reported here shows some variation from the typical multicopy cyclotide precursor. Unlike other cyclotide precursors, the entire region between the signal sequence and the mature tricyclon A sequence is repeated prior to the sequence for tricyclon B. Effectively, the tricyclon A NTR is greatly expanded in size by the incorporation of the proregion that is present between the signal sequence and the initial repeated domain in other cyclotide precursors. The structures of the NTRs, determined as discrete peptide fragments rather than as part of a cyclotide precursor, from *O. affinis* and

V. odorata adopt amphipatic helices, and it has been speculated that they may be involved in stabilizing the surface-exposed hydrophobic patch during processing (Dutton et al., 2004). Structure prediction indicates that the tricyclon A NTR may adopt two helical sections (Figure 6). The short proregions of the *Oak*, *Voc*, and *Vok* clones are also predicted to adopt a helical structure; hence, the predicted helices in the tricyclon A NTR may correspond to the typical cyclotide NTR and the proregion that is not repeated in other cyclotides but forms part of the NTR in tricyclon A. The expansion of the NTR in the tricyclon clone may indicate that the precursor structure is amenable to variation without affecting the processing mechanism. One way in which this could be achieved is by digestion of the precursor into discrete units prior to cyclization. The greater size, and presumed structural perturbation, caused by the NTR may also reflect the fact that the lower surface hydrophobicity of tricyclon A negates the need for stabilization by an amphipatic helix.

Conclusions

Gene-expressed cyclic proteins were unknown until recently, and apart from an interesting biological curio, they offer exciting potential for the development of biologically active and stable protein drugs. However, little is known about the biosynthesis and evolution of macrocyclic proteins in general and cyclotides in particular. Resolution of these questions offers the opportunity to develop technologies that could be applied to other proteins with biologically important activities. The structure determined here for tricyclon A shows that cyclotides may be engineered to reduce hemolytic activity and may be able to incorporate novel structural motifs.

Experimental Procedures

High-Pressure Liquid Chromatography

Preparative reverse phase-high-pressure liquid chromatography (RP-HPLC) was performed on a Waters 600-MS System Controller system equipped with a Waters 484 Tunable Absorbance Detector. Samples were manually loaded onto a Vydac C18 column (22 × 250 mm, 300 Å pore size, 10 μm particle size) and eluted at a flow rate of 8 ml/min with a linear gradient of 0%–80% buffer B (90% HPLC grade acetonitrile in H_2O/0.09% trifluoroacetic acid) for 80 min with UV detection at 230 nm. Semipreparative RP-HPLC was performed by using the same conditions and equipment with a Vydac C18 column (10 × 250 mm, 300 Å pore size, 10 μm particle size) and a flow rate of 3 ml/min. Analytical RP-HPLC was performed on a Waters 600 Controller system equipped with a Waters 486 Tunable Absorbance Detector. Samples were auto-injected by using a Waters 717plus onto a Vydac C18 analytical column (4.6 × 250 mm, 300 Å pore size, 5 μm particle size). Samples were eluted at a flow rate of 1 ml/min with a linear gradient of 0%–80% buffer B for 40 min with UV detection at 215 nm.

Mass Spectrometry

All mass data were obtained by using electrospray/time-of-flight mass spectrometry on a Mariner instrument (PerSeptive Biosystems). A 25 μl sample was injected into an Applied Biosystems 140B Solvent Delivery System and passed through the mass spectrometer at a flow rate of 80 μl/min with a constant organic solvent ratio of 80% buffer B. Protein mass peaks were recorded in the linear mode between 1,000 and 25,000 m/z^+ with a step size of 0.1 or 0.2 Da and a delay time of 0.3 s. The nozzle potential was set between 60 and 100 V, and data were acquired as $[M+2H]^2$ and processed by using Biospec Data Explorer 3.0.0.0 (PerSeptive Biosystems) software.

Extraction and Isolation of Tricyclon A

Flowers from *V. tricolor*, locally grown in Brisbane, were pulped in methanol by using a kitchen blender. Plant debris was removed by filtering through a wire strainer. Samples were extracted at room temperature in an equal volume of dichloromethane for 4 hr. Remaining plant debris was removed by filtering through a cotton plug filter, and the filtrate was repeatedly (4–5 times) partitioned in chloroform and water. The water phase was collected, and the remaining methanol was evaporated in vacuo. The resulting solution was lyophilized to obtain the dry plant extract.

Approximately 20–300 mg dry plant extract was dissolved in 10 ml 50% buffer B and loaded onto the preparative HPLC system. Peaks containing tricyclon A were collected in glass test tubes, and the presence of proteins was confirmed with mass spectrometry. These samples were lyophilized and further purified by using semipreparative RP-HPLC. Peaks were collected in glass test tubes, and the constituents of the peaks were determined as described above. The peak containing tricyclon A was lyophilized to obtain the dry protein extract.

Reduction and Alkylation Reactions

Reduction was performed with a protein concentration of 100 μg/ml in a 0.1 M Tris-HCl buffer (pH 8.0) with the addition of guanidine-HCl to a concentration of 6 M. A 10-fold molar excess of DDT was added, and the reaction was incubated for 5 hr at 37°C. Cys residues were alkylated by the addition of 0.2 M iodoacetamide, 6 M guanidine-HCl, and 0.5 M Tris-HCl (pH 8.0) to the reduction reaction to give a 5-fold molar excess of iodoacetamide over total thiols in the reaction. The reaction was incubated for 12 hr. Light was excluded to minimize the risk of iodine formation from available iodide ions.

NMR Spectroscopy and Structure Calculations

Approximately 2 mg dry tricyclon A was dissolved in 550 μl 90% H_2O/10% D_2O (99.9%, Cambridge Isotope Laboratories, Woburn, MA), with 100 μl CD_3CN (99.8% Cambridge Isotope Laboratories, Woburn, MA) added to increase solubility; the pH of the sample was measured as 3.0. Spectra were recorded at 305 and 295 K on a Bruker DMX-750 NMR spectrometer. All two-dimensional (2D) spectra were recorded in phase-sensitive mode by using time-proportional phase incrementation for quadrature detection in the t1 dimension (Marion and Wuthrich, 1983). 2D experiments performed included a DQF-COSY (Rance et al., 1983), a TOCSY (Braunschweiler and Ernst, 1983) with a mixing time of 80 ms, an ECOSY (Griesinger et al., 1987) in D_2O, and a NOESY spectra (Jeener et al., 1979) with mixing times of 100, 150, and 200 ms. Solvent suppression for NOESY and TOCSY experiments was achieved by using a modified WATERGATE sequence (Piotto et al., 1992). For the DQF-COSY experiment, the water signal was suppressed by lower power irradiation during the relaxation delay (1.8 s). Spectra were routinely acquired over 8802 Hz with 4096 complex data points in F2 and 512 increments in the F1 dimension, with 40 scans per increment (96 for NOESY). A series of one-dimensional (1D) and TOCSY spectra was run immediately after dissolving the fully protonated sample in D_2O for identification of slowly exchanging amide protons.

Spectra were processed on a Silicon Graphics Indigo workstation by using XWIN-NMR (Bruker) software. The t1 dimension was zero filled to 2048 data points, and shifted sinebell window functions were applied prior to Fourier transformation. Chemical shifts were referenced to DSS at 0.00 ppm. Spectra were analyzed with the program SPARKY (Goddard and Kneller, 2005). The intensities of cross-peaks in NOESY spectra with a mixing time of 100 ms were calibrated, and distance constraints were derived by using CYANA (Guntert et al., 1997). Corrections for pseudoatoms were added to distance constraints where needed (Wuthrich et al., 1983). Backbone dihedral angle restraints were derived from $^3J_{HNH\alpha}$ coupling constants measured from line-shape analysis of antiphase cross-peak splitting in the DQF-COSY spectrum. Angles were restrained to $-120° \pm 40°$ for $^3J_{HNH\alpha} > 9$ Hz . After initial structure calculations with CYANA, sets of 50 structures were calculated by using a torsion angle-simulated annealing protocol within CNS (Brunger et al., 1997). Each set of structures was then subjected to further molecular dynamics and energy minimization in a water

shell (Linge and Nilges, 1999) before further refinement. During refinement with CNS, positions of slowly exchanging amides were analyzed, and in 12 cases, hydrogen bonds were unambiguously assigned and added to the restraints. The force constants used in the final rMD/rEM were 50 kcal $mol^{-1}Å^{-1}$ for distant restraints and 200 kcal $mol^{-1}deg^{-1}$ for angle restraints.

Protein Sequencing

The sequence of tricyclon A was determined by using a combination of Edman and MS-MS sequencing. For Edman sequencing, dry reduced/alkylated tricyclon A (1 mg) was resuspended in 50 μl NH_4HCO_3 (pH 8.0), and 2 μl 1.5 mg/ml trypsin in NH_4HCO_3 (pH 8.0) was added. The sample was incubated at 37°C for 3 hr. Analytical RP-HPLC was performed, peaks were collected, and the peak constituents were identified by their mass. Automated Edman degradation yielded a protein sequence of 23 residues.

For MS-MS sequencing, the fragments resulting from tryptic digests were sequenced by MS/MS on an API QStar mass spectrometer (PE Sciex) equipped with a nanospray ionization source (MDS Proteomics A/S, Odense, Denmark). All samples were sprayed in 50% acetonitrile, 0.5% formic acid in water. Typically, 2 μl was loaded into the nanospray needle. A capillary voltage of 900V was applied, and the instrument was operated in positive ion mode. Data were acquired between a mass range of 60–2000. Data were collected, processed, and analyzed by using the Analyst QS Software package (Applied Biosystems/MDS Sciex, 2001).

RNA Extraction, RT-PCR, RACE, and Secondary Structure Prediction

RNA was isolated by using the RNAqueous Kit from Ambion, Inc. Single-stranded cDNA was prepared from leaf RNA by using the OneStep RT-PCR Kit from Qiagen Pty. Ltd. A partial clone was amplified by the PCR by using a degenerate primer coding for the sequence CGEWK (5′-TG[CT]GGIGA[AG]TGGA-3′) and oligo-dT (Proligo Australia Pty. Ltd., Lismore, NSW, Australia). The resulting band was excised from agarose gel and cloned into pCR 2.1-TOPO vector by using Invitrogens TOPO Cloning Kit for sequencing. A 5′-RACE library was constructed by using the FirstChoice RLM-RACE Kit from Ambion, Inc. Clones were obtained by screening this library by using the PCR with gene-specific primers derived from the sequencing of the RT-PCR clone (5′-GCTGCACGAGTGATGA CATCTTT-3′) coding for the sequence KDVITRA (Proligo Australia Pty. Ltd.) and the kit-supplied 5′ primer. The resulting bands were excised from agarose gel, cloned, and sequenced as described above. The amino acid sequence corresponding to the NTR of the identified precursor was analyzed for secondary structure by using the PHDsec (Rost and Sander, 1993) server at http://www.embl-heidelberg.de/predictprotein/submit_def.html.

Hemolytic Assays

An approximately 500 μM solution of tricyclon A in DMSO was serially diluted in phosphate-buffered saline (PBS) to give 20 μl test solutions in a 96-well, U-bottomed microtiter plate (Nunc). Human type A red blood cells (RBCs) were washed with PBS and centrifuged at 4000 rpm for 30 s in a microcentrifuge several times until a clear supernatant was obtained. A 0.25% suspension of washed RBCs in PBS was added (100 μl) to the protein solutions. The plate was incubated at 37°C for 1 hr and centrifuged at 150 × g for 5 min. Aliquots of 100 μl were transferred to a 96-well, flat-bottomed microtiter plate (Falcon), and the absorbency was measured at 405 nm with an automatic Multiskan Ascent plate reader (Labsystems). The level of hemolysis was calculated as the percentage of maximum lysis (1% Triton X-100 control) after adjusting for minimum lysis (PBS control). Synthetic melittin (Sigma) was used for comparison.

Acknowledgments

This work was funded by a grant from the Australian Research Council (ARC). We thank Norelle Daly and Johan Rosengren for help with structure calculations and Mariel Felizmenio-Quimio for providing a comparative sample of tricyclon A for analysis. J.P.M. is the recipient of an Australian Research Council postgraduate award.

Received: November 18, 2004
Revised: January 30, 2005
Accepted: February 3, 2005
Published: May 10, 2005

References

Barry, D.G., Daly, N.L., Clark, R.J., Sando, L., and Craik, D.J. (2003). Linearization of a naturally occurring circular protein maintains structure but eliminates hemolytic activity. Biochemistry *42*, 6688–6695.

Barry, D.G., Daly, N.L., Bokesch, H.R., Gustafson, K.R., and Craik, D.J. (2004). Solution structure of the cyclotide palicourein: implications for the development of a pharmaceutical framework. Structure *12*, 85–94.

Braunschweiler, L., and Ernst, R.R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. J. Magn. Reson. *53*, 521–528.

Brown, J.C., Jolliffe, N.A., Frigerio, L., and Roberts, L.M. (2003). Sequence-specific, Golgi-dependent vacuolar targeting of castor bean 2S albumin. Plant J. *36*, 711–719.

Brunger, A.T., Adams, P.D., and Rice, L.M. (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. Structure *5*, 325–336.

Bulet, P., Stocklin, R., and Menin, L. (2004). Anti-microbial peptides: from invertebrates to vertebrates. Immunol. Rev. *198*, 169–184.

Camarero, J.A., Fushman, D., Sato, S., Giriat, I., Cowburn, D., Raleigh, D.P., and Muir, T.W. (2001). Rescuing a destabilized protein fold through backbone cyclization. J. Mol. Biol. *308*, 1045–1062.

Colgrave, M.L., and Craik, D.J. (2004). Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. Biochemistry *43*, 5965–5975.

Craik, D.J. (2001). Plant cyclotides: circular, knotted peptide toxins. Toxicon *39*, 1809–1813.

Craik, D.J., Daly, N.L., Bond, T., and Waine, C. (1999). Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J. Mol. Biol. *294*, 1327–1336.

Craik, D.J., Daly, N.L., and Waine, C. (2001). The cystine knot motif in toxins and implications for drug design. Toxicon *39*, 43–60.

Craik, D.J., Simonsen, S., and Daly, N.L. (2002). The cyclotides: novel macrocyclic peptides as scaffolds in drug design. Curr. Opin. Drug Discov. Devel. *5*, 251–260.

Craik, D.J., Daly, N.L., Mulvenna, J., Plan, M.R., and Trabi, M. (2004). Discovery, structure and biological activities of the cyclotides. Curr. Protein Pept. Sci. *5*, 297–315.

Daly, N.L., Koltay, A., Gustafson, K.R., Boyd, M.R., Casas-Finet, J.R., and Craik, D.J. (1999a). Solution structure by NMR of circulin A: a macrocyclic knotted peptide having anti-HIV activity. J. Mol. Biol. *285*, 333–345.

Daly, N.L., Love, S., Alewood, P.F., and Craik, D.J. (1999b). Chemical synthesis and folding pathways of large cyclic polypeptides: studies of the cystine knot polypeptide kalata B1. Biochemistry *38*, 10606–10614.

Deechongkit, S., and Kelly, J.W. (2002). The effect of backbone cyclization on the thermodynamics of β-sheet unfolding: stability optimization of the PIN WW domain. J. Am. Chem. Soc. *124*, 4980–4986.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (http://www.pymol.org).

Dombrowski, J.E., Schroeder, M.R., Bednarek, S.Y., and Raikhel, N.V. (1993). Determination of the functional elements within the vacuolar targeting signal of barley lectin. Plant Cell *5*, 587–596.

Dou, Y., Baisnee, P.F., Pollastri, G., Pecout, Y., Nowick, J., and Baldi, P. (2004). ICBS: a database of interactions between protein chains mediated by β-sheet formation. Bioinformatics *20*, 2767–2777.

Dutton, J.L., Renda, R.F., Waine, C., Clark, R.J., Daly, N.L., Jennings, C.V., Anderson, M.A., and Craik, D.J. (2004). Conserved structural and sequence elements implicated in the processing of gene-encoded circular proteins. J. Biol. Chem. *279*, 46858–46867.

Frigerio, L., de Virgilio, M., Prada, A., Faoro, F., and Vitale, A. (1998). Sorting of phaseolin to the vacuole is saturable and requires a short C-terminal peptide. Plant Cell *10*, 1031–1042.

Frigerio, L., Jolliffe, N.A., Di Cola, A., Felipe, D.H., Paris, N., Neuhaus, J.M., Lord, J.M., Ceriotti, A., and Roberts, L.M. (2001). The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting. Plant Physiol. *126*, 167–175.

Goddard, T.D., and Kneller, D.G. (2005). SPARKY 3 (computer program). University of California, San Francisco.

Griesinger, C., Sorensen, O.W., and Ernst, R.R. (1987). Practical aspects of the E-COSY technique, measurement of scalar spin-spin coupling constants in peptides. J. Magn. Reson. *75*, 474–492.

Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.

Gustafson, K.R., Sowder, R.C., Henderson, L.E., Parsons, I.C., Kashman, Y., Cardellina, J.H., McMahon, J.B., Buckheit, R.W., Pannell, L.K., and Boyd, M.R. (1994). Circulins A and B: novel HIV-inhibitory macrocyclic peptides from the tropical tree *Chassalia parvifolia*. J. Am. Chem. Soc. *116*, 9337–9338.

Hutchinson, E.G., and Thornton, J.M. (1996). PROMOTIF: a program to identify and analyze structural motifs in proteins. Protein Sci. *5*, 212–220.

Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1979). Investigation of exchange processes by two dimensional NMR spectroscopy. J. Chem. Phys. *71*, 4546–4553.

Jennings, C., West, J., Waine, C., Craik, D., and Anderson, M. (2001). Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. Proc. Natl. Acad. Sci. USA *98*, 10614–10619.

Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55.

Linge, J.P., and Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. J. Biomol. NMR *13*, 51–59.

Marion, D., and Wuthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H–1H spin-spin coupling constants in proteins. Biochem. Biophys. Res. Commun. *113*, 967–974.

Martin, L., Stricher, F., Misse, D., Sironi, F., Pugniere, M., Barth, P., Prado-Gotor, R., Freulon, I., Magne, X., Roumestand, C., et al. (2003). Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. Nat. Biotechnol. *21*, 71–76.

Matsuoka, K., and Nakamura, K. (1999). Large alkyl side-chains of isoleucine and leucine in the NPIRL region constitute the core of the vacuolar sorting determinant of sporamin precursor. Plant Mol. Biol. *41*, 825–835.

Muntz, K., and Shutov, A.D. (2002). Legumains and their functions in plants. Trends Plant Sci. *7*, 340–344.

Neuhaus, J.M., Pietrzak, M., and Boller, T. (1994). Mutation analysis of the C-terminal vacuolar targeting peptide of tobacco chitinase: low specificity of the sorting system, and gradual transition between intracellular retention and secretion into the extracellular space. Plant J. *5*, 45–54.

Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J. Biomol. NMR *2*, 661–665.

Rance, M., Sorensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wuthrich, K. (1983). Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering. Biochem. Biophys. Res. Commun. *27*, 157–162.

Rosengren, K.J., Daly, N.L., Plan, M.R., Waine, C., and Craik, D.J. (2003). Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. J. Biol. Chem. *278*, 8606–8616.

Rost, B., and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. *232*, 584–599.

Saalbach, G., Rosso, M., and Schumann, U. (1996). The vacuolar targeting signal of the 2S albumin from Brazil nut resides at the C terminus and involves the C-terminal propeptide as an essential element. Plant Physiol. *112*, 975–985.

Saether, O., Craik, D.J., Campbell, I.D., Sletten, K., Juul, J., and Norman, D.G. (1995). Elucidation of the primary and three-dimensional structure of the utertonic polypeptide kalata B1. Biochemistry *34*, 4147–4158.

Schopke, T., Hasan Agha, M.I., Kraft, R., Otto, A., and Hiller, K. (1993). H ämolytisch aktive komponenten aus Viola tricolor l. und Viola arvensis. Murray. Sci. Pharm. *61*, 145–153.

Tam, J.P., Lu, Y.A., Yang, J.L., and Chiu, K.W. (1999). An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. Proc. Natl. Acad. Sci. USA *96*, 8913–8918.

Tedford, H.W., Fletcher, J.I., and King, G.F. (2001). Functional significance of the β-hairpin in the insecticidal neurotoxin ω-Atracotoxin-hv1a. J. Biol. Chem. *276*, 26568–26576.

Trabi, M., and Craik, D.J. (2002). Circular proteins — no end in sight. Trends Biochem. Sci. *27*, 132–138.

Trabi, M., and Craik, D.J. (2004). Tissue-specific expression of head-to-tail cyclized miniproteins in Violaceae and structure determination of the root cyclotide *Viola hederacea* root cyclotide1. Plant Cell *16*, 2204–2216.

Witherup, K.M., Bogusky, M.J., Anderson, P.S., Ramjit, H., Ransom, R.W., Wood, T., and Sardana, M. (1994). Cyclopsychotride A, a biologically active, 31-residue cyclic peptide isolated from *Psychotria longipes*. J. Nat. Prod. *57*, 1619–1625.

Wuthrich, K., Billeter, M., and Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J. Mol. Biol. *169*, 949–961.

Accession Numbers

The structure reported in this paper was deposited in the PDB and has the code 1YP8.